Introduction to Bioinformatics
1. **Databases and Literature Searches**

Benjamin F. Matthews
United States Department of Agriculture
Soybean Genomics and Improvement Laboratory
Beltsville, MD 20708
matthewb@ba.ars.usda.gov

Outline of course

- Introduction to Bioinformatics
- Databases
- Literature searches
- DNA sequence retrieval
- DNA analysis
  - Translation
  - alignment
- Protein retrieval and analysis
- Phylogenetic trees
- Restriction analysis
- Primer design
- Genomic sequencing and annotation
- Networks and pathways
- Gene expression

What we will cover today

- What is bioinformatics
- Databases
- Literature searches

Bioinformatics

-the use of computers to collect, analyze, and interpret biological information

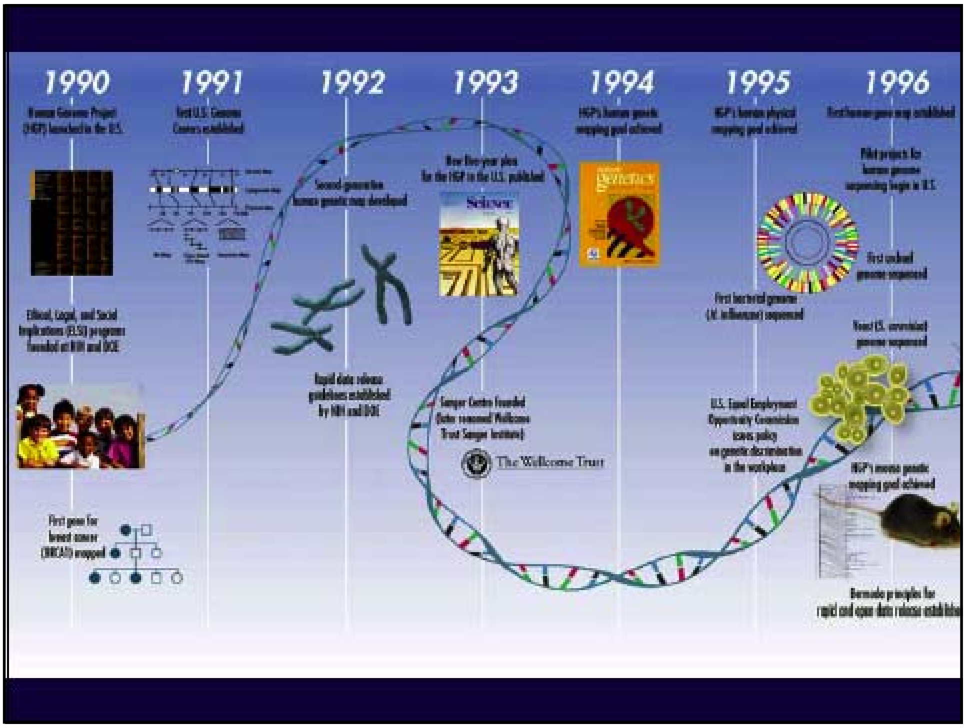
-a set of software tools and a database

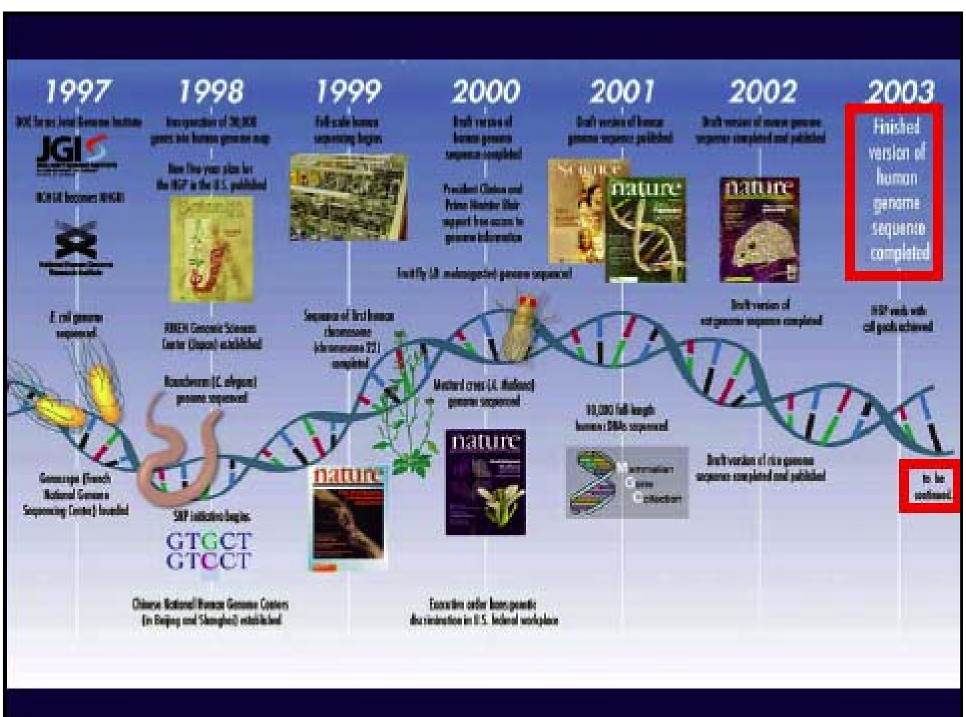
Genomics Revolution

- Costs of DNA sequencing and novel techniques (eg. microarrays) have fallen dramatically
- Hugh amount of genomics sequence is available
- Novel techniques are more readily available
- Huge amounts of data is moving biology toward becoming an information science

New challenge for biologists

- Organize and mine data
- Identify genes
- Determine gene and protein function
- Identify important and useful genes
  - Diagnostic
  - Modification
  - New value-added traits


1990
Human Genome Project (HGP) launched in the U.S.
Ethical, Legal, and Social Implications (ELSI) programs founded at NIH and DOE
First gene for breast cancer (BRCA1) mapped
1991
First U.S. Genome Centers established
1992
Second-generation human genetic map developed
Rapid data release guidelines established by NIH and DOE
1993
New five-year plan for the HGP in the U.S. published
Sanger Centre founded (later renamed Wellcome Trust Sanger Institute)
The Wellcome Trust
1994
HGP's human genetic mapping goal achieved
1995
HGP's human physical mapping goal achieved
First bacterial genome (H. influenzae) sequenced
U.S. Equal Employment Opportunity Commission issues policy on genetic discrimination in the workplace
1996
First human gene map established
Pilot projects for human genome sequencing begin in U.S.
First archaeal genome sequenced
Yeast (S. cerevisiae) genome sequenced
HGP's mouse genetic mapping goal achieved
Bermuda principles for rapid and open data release established


1997
DOE forms Joint Genome Institute
JGI
NCHGR becomes NHGRI
E. coli genome sequenced
Genoscope (French National Genome Sequencing Center) founded
1998
Incorporation of 30,000 genes into human genome map
New five-year plan for the HGP in the U.S. published
RIKEN Genomic Sciences Center (Japan) established
Roundworm (C. elegans) genome sequenced
SNP initiative begins
GTGCT
GTCCT
Chinese National Human Genome Centers (in Beijing and Shanghai) established
1999
Full-scale human sequencing begins
Sequence of first human chromosome (chromosome 22) completed
nature
2000
Draft version of human genome sequence completed
President Clinton and Prime Minister Blair support free access to genome information
Fruitfly (D. melanogaster) genome sequenced
Mustard cress (A. thaliana) genome sequenced
nature
Executive order bans genetic discrimination in U.S. federal workplace
2001
Draft version of human genome sequence published
Science
nature
10,000 full-length human cDNAs sequenced
Mammalian Gene Collection
2002
Draft version of mouse genome sequence completed and published
nature
Draft version of rat genome sequence completed
Draft version of rice genome sequence completed and published
2003
Finished version of human genome sequence completed
HGP ends with all goals achieved
to be continued.

So much data!

- Need to store data in organized fashion
- Need rapid retrieval of portions of data
- Need to compare data
- Need to annotate data

National Center for Biotechnology Information

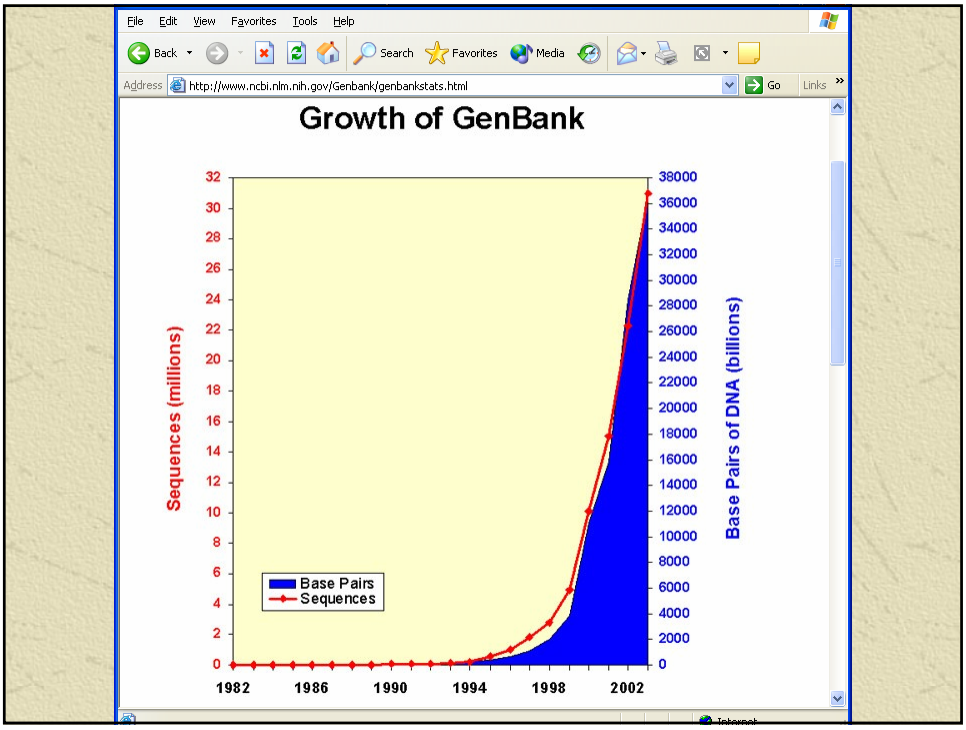
http://www.ncbi.nlm.nih.gov/

Created as a part of NLM in 1988

- To establish public databases
  - U.S. National DNA Sequence Database
- To perform research in computational biology
- To develop software tools for sequence analysis
- To disseminate biomedical information

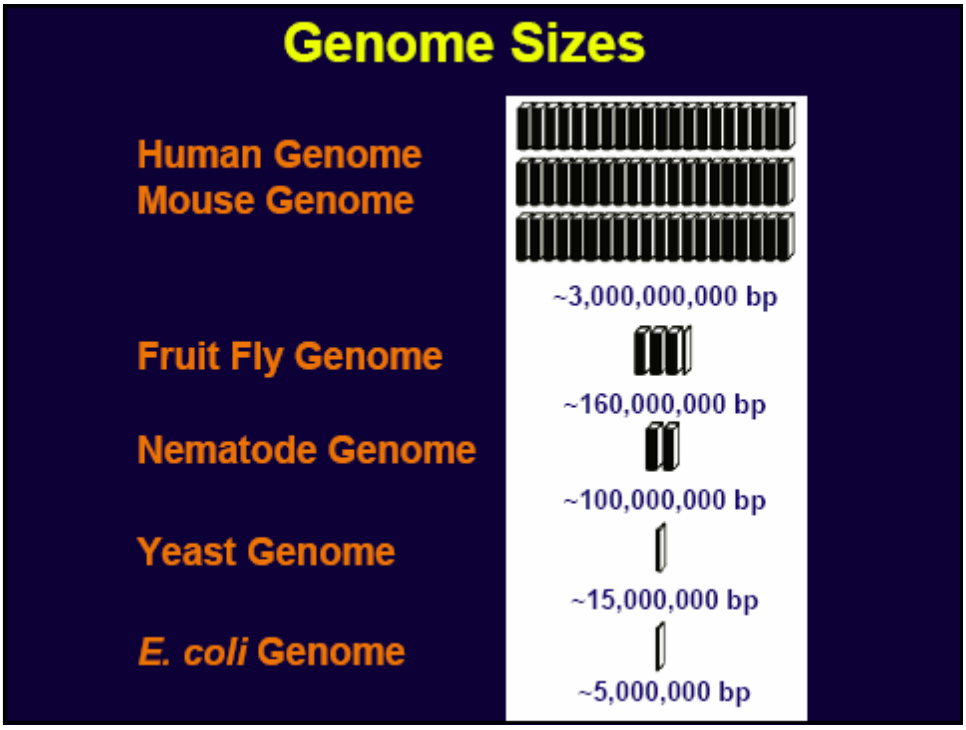
GenBank

- http://www.ncbi.nlm.nih.gov/Genbank/index.html
- Nucleotide sequences
- >130,000 organisms
- Annotated records with coding region features and amino acid translations



Genome update

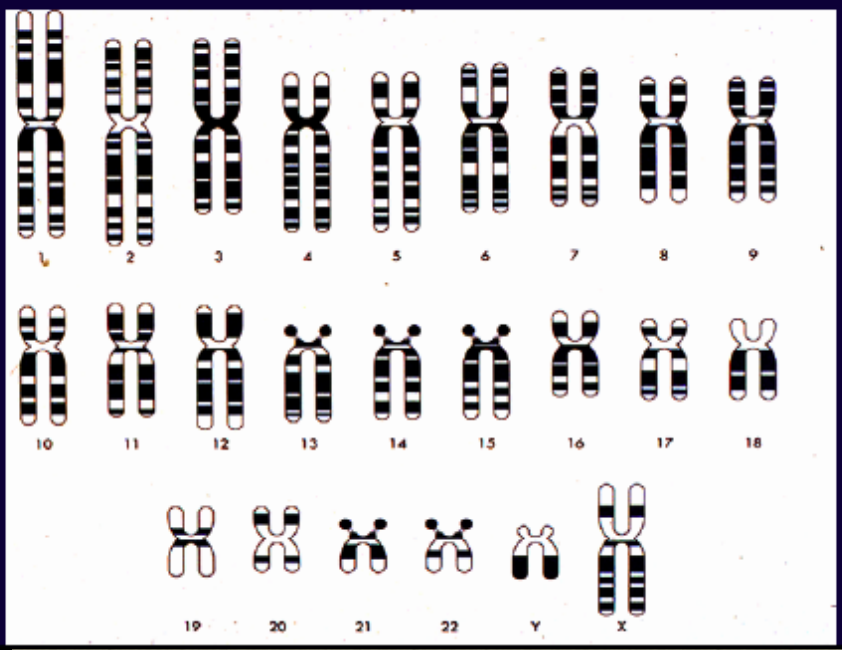
- Numerous genome drafts
  - 155 bacteria
  - 18 archaea
  - numerous viruses
  - Fungi (12 or more)
  - Animal
    - Caenorhabditis elegans, drosophila, mosquito, honey bee, zebra fish, chicken, mouse, rat, dog, chimpanzee, human
  - Plant
    - Arabidopsis, Rice
- Many more under construction

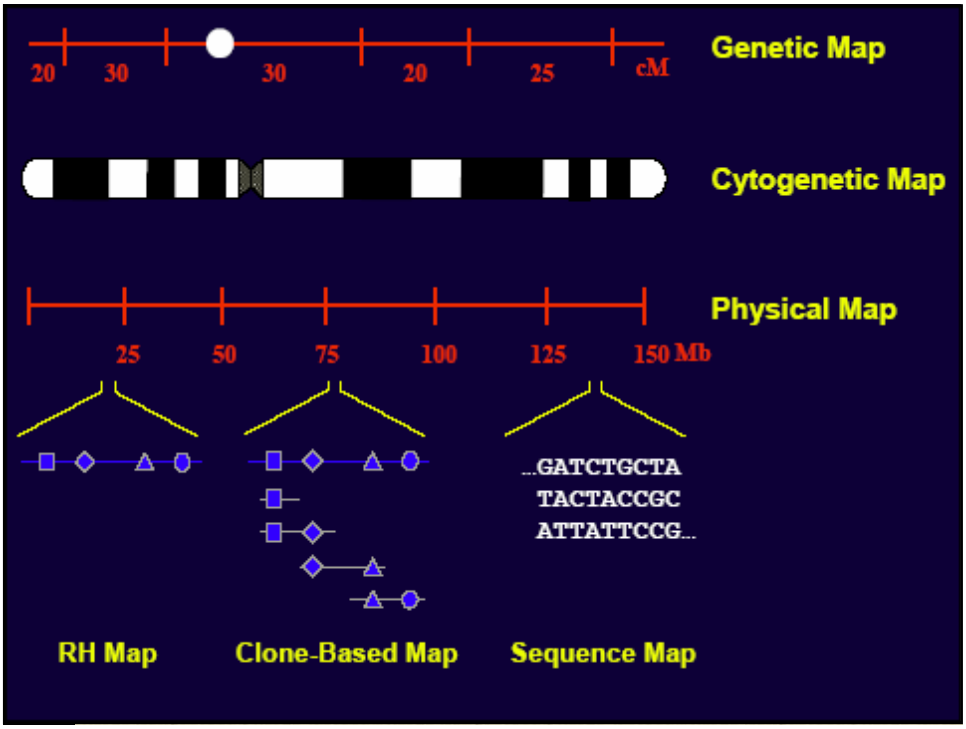


Human Genome

- 32 billion base pairs
- Close to a complete list of human genes and proteins
- Challenge is to understand the genes, proteins and biological processes, so we can use this information to help the world

The Human Cytogenetic Map





Curriculum for bioinformatics students

- Biology
  - Molecular biology, biochemistry, physiology
- Statistics, mathematics
  - Data analysis, Algorithm development
- Computer programming
  - Software design and development
  - Perl, C++
- Database design and management
  - SQL
- Web site design and management
- Use of current bioinformatics software packages and databases

A biology student

- Strong biology background
- Use of existing databases
- Use of existing software

- Does not necessarily need programming language

Databases

A Database

- Self-describing collection of integrated records
- Computerized
- Complete with language for access

- There are many databases and software tools available to mine the databases
- Provide permanent storage of information
- Application programs allow users to manipulate information

Biological Database

- A large, organized body of persistent data
- Associated with computerized software
  - Analyze
  - Update
  - Query
  - Retrieve
- Easy access to information
- Method of extraction of needed information only

Record

- a representation of some physical or conceptual object
- Eg. A record of business customers
  - A record contains **attributes**
  - Eg. Name, address, telephone number
- An individual name, address, telphone number is **data**

Database Management System (DBMS)

- A set of programs used to define, administer, and process databases

Conventional Operating Systems (Hierarchical)

- Information was placed in a record and stored in a file system
- Application programs were written to access information as needs arose
- As needs change, new permanent files must be created and new application programs must be written

Relational Database

- Can change database structure without rewriting applications
- Data may be retrieved from various files and from various formats
- Reduce redundancy and inconsistency
  - Reduces storage and access costs
- Easier access of information
  - New application programs need not be written
- Easier to enforce consistency constraints
- Easier to enforce security constraints
- Therefore it is more flexible

Some Biological Databases

- Literature (PubMed and Agricola)
  - Journals
  - books
- DNA (GenBank)
  - cDNA and EST
  - Whole genome
- Protein (Etrez Protein and ExPASy)
  - Sequence
  - Structure
- Metabolic (KEGG regulatory pathways)
- Organism
  - Human, mouse, drosophila, Arabidopsis

Major Molecular Databases

- NCBI (National Center for Biotechnology Information): http://www.ncbi.nih.gov
  - Contains GenBank, PubMed, other databases, tools & links
- GenBank: http://www.ncbi.nih.gov/Genbank/
- EMBL (European Molecular Biology Laboratory) http://www.ebi.ac.uk/embl/
- PDB (Protein Data Bank):
  - http://www.rcsb.org/pdb/

Major Molecular Databases

- PIR (Protein Information Resource):
  - http://pir.georgetown.edu/
- ExPasy proteomics server (Expert Protein Analysis System):
  - http://us.expasy.org/
- KEGG (Kyoto Encyclopedia of Genes and Genomes):
  - http://www.genome.ad.jp/kegg/kegg2.html

National Center for Biotechnology Information

http://www.ncbi.nlm.nih.gov/

Created as a part of NLM in 1988

- To establish public databases
  - U.S. National DNA Sequence Database
- To perform research in computational biology
- To develop software tools for sequence analysis
- To disseminate biomedical information

NCBI houses several databases

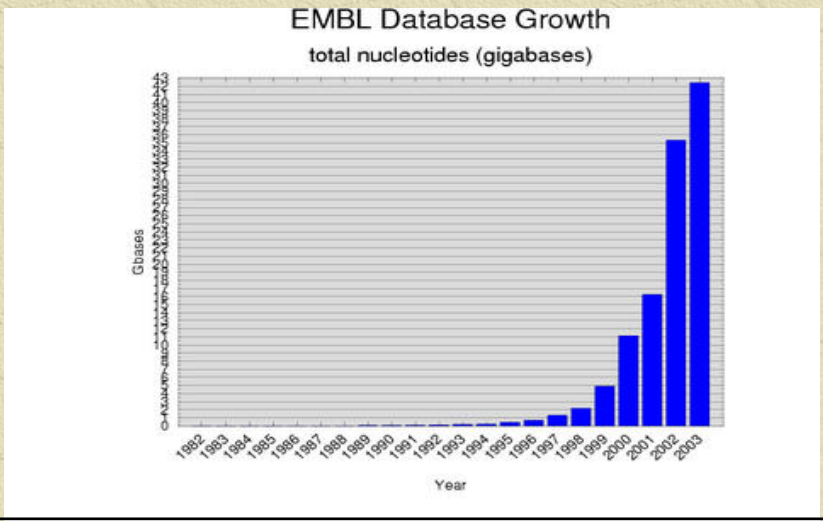
- Entrez
- PubMed
- Nucleotide
- Protein
- Genome
- Structure
- Taxonomy

http://www.ncbi.nlm.nih.gov/

EMBL Nucleotide Sequence Database

- Website: http://www.ebi.ac.uk/embl/
- Maintained at Europe Institute of Bioinformatics and is Europe's primary nucleotide sequence resource
- Include both DNA and RNA sequences
- From direct submissions of individual researchers, genome sequencing projects and patent applications
- Daily exchanges with GenBank and DDBJ

EMBL Database Statistics



ExPasy

- Abbreviated from Expert Protein analysis System
- Web site: http://us.expasy.org/
- The ExPASy server is maintained by the Swiss Institute of Bioinformatics
- Contains molecular databases, tools, and links

KEGG

- GeneUniverse
  - Genes and gene products in complete genomes
- Chemical networks
  - Chemical compounds
  - Chemical reactions
- Protein Networks
  - Metabolic pathways
  - Regulatory pathways
  - Relationships

Databases

- Organize data
- Allows data submission and access
- Resource for other databases and tools
- Find what you want…fast

Database components

- Define and describe
- Unique identifier
- Update version
- Links to other databases
- Documentation
- Submission/update/correction process
- Retrieval system
- Standards
  - Universal if appropriate

Retrieval system

- User interface
  - Computer
- Batch mode
  - Avoid repetitive process
- Structured queries or SQL (Structured Query Language) access
  - Allows mining of data
  - User-friendlydata mining
- Full-dump
  - Allows placement of data on other computers
- All data accessible

Retrieval System (cont.)

- Documentation
  - How things work
- Link definitions
  - What is at each spot
- User support
  - If something goes wrong

Cost

- Hardware-computer, screen, keyboard, mouse, extra storage
- Software
- Production cost
  - Training
  - data accumulation and entry
- Usage cost
  - Computer, software, training, internet access

Literature databases and literature searches

You Define A Problem

- Target problem area
  - Meet with farmers, extension agents, colleagues
  - Read agricultural reports and literature
  - Attend meetings
- Study problem
  - Read literature
  - Talk with colleagues
  - Go to meetings and seminars

Define Project Goals

- Intensive, focused literature search
- Where can you conduct a literature search on-line?

Literature searches

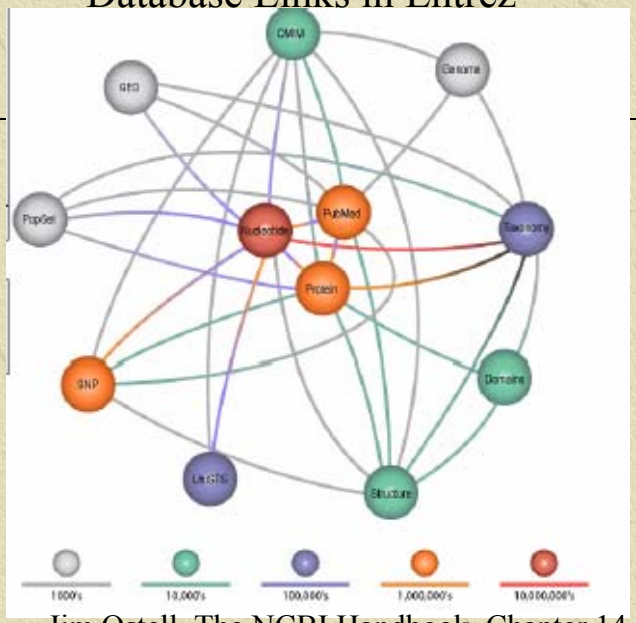
- NCBI Pub Med
  - http://www.ncbi.nlm.nih.gov/PubMed/
- USDA Agricola
  - http://agricola.nal.usda.gov/

National Center for Biotechnology Information

- Established 1988
- Creates
  - Public databases
  - Conducts research in computational biology
  - Develops software
  - Disseminates biomedical information
  - Entrez
    - Search and retrieval from many databases

http://www.ncbi.nlm.nih.gov/Entrez/

Database Links in Entrez



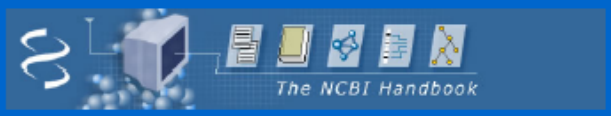
Jim Ostell, The NCBI Handbook, Chapter 14

Entrez: Search and Retrieval System for NCBI Databases

Organism	Phylogenetic and taxonomic information
Nucleotide	GenBank, EMBL, DDBJ, RefSeq, PDB
PopSet	Population study datasets
Genome	Complete genomes
SNP	Single nucleotide polymorphism
EST	Expressed sequence tags
UniGene	Clusters of expressed sequences
GEO	Microarray datasets
Protein	Translations of GenBank & RefSeq records, SWISS-PROT, PIR, PRF, PDB
Domains	CDD: conserved domain database
Structure	MMDB:experimental 3D structures
Pubmed	Biomedical literature
PubMed Central	Free online journals
Books	Free online textbooks

NCBI

Online Books

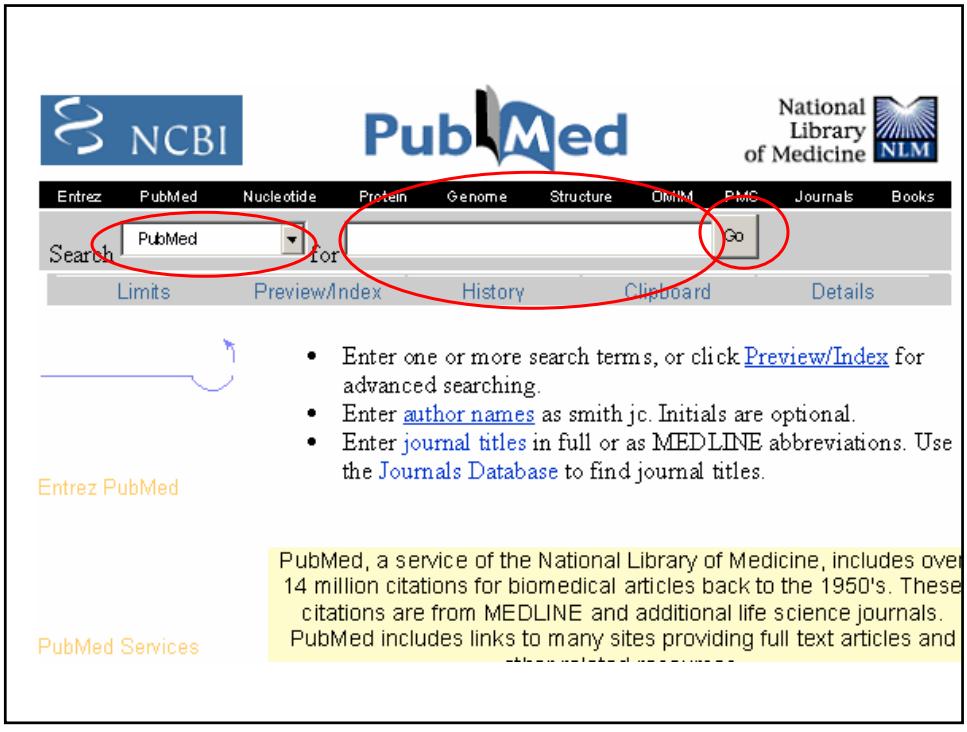


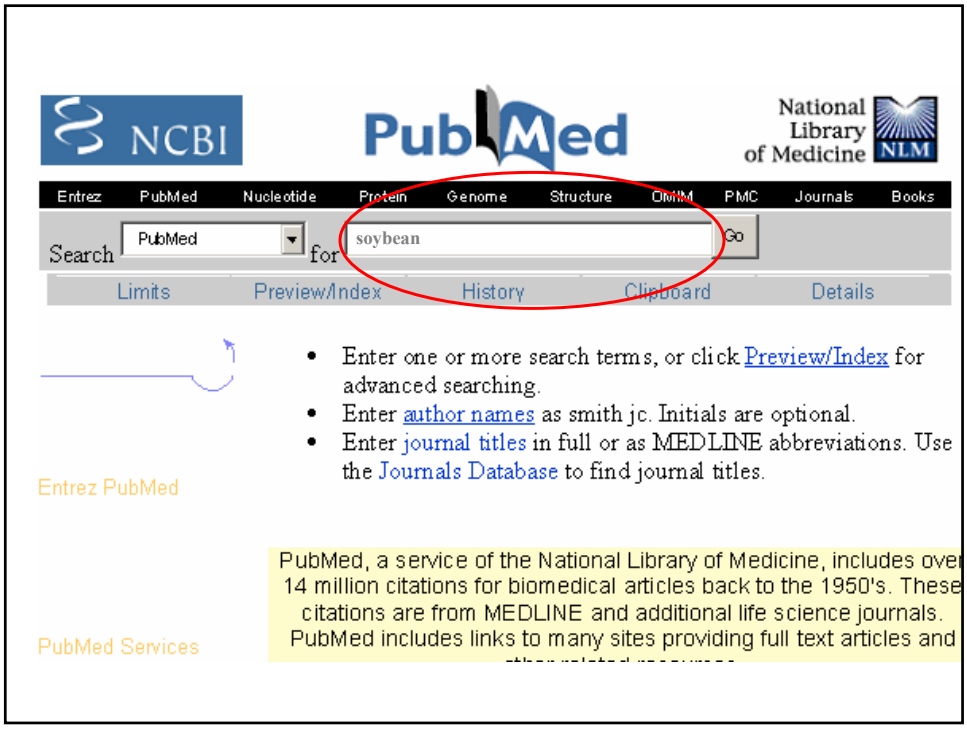
NCBI

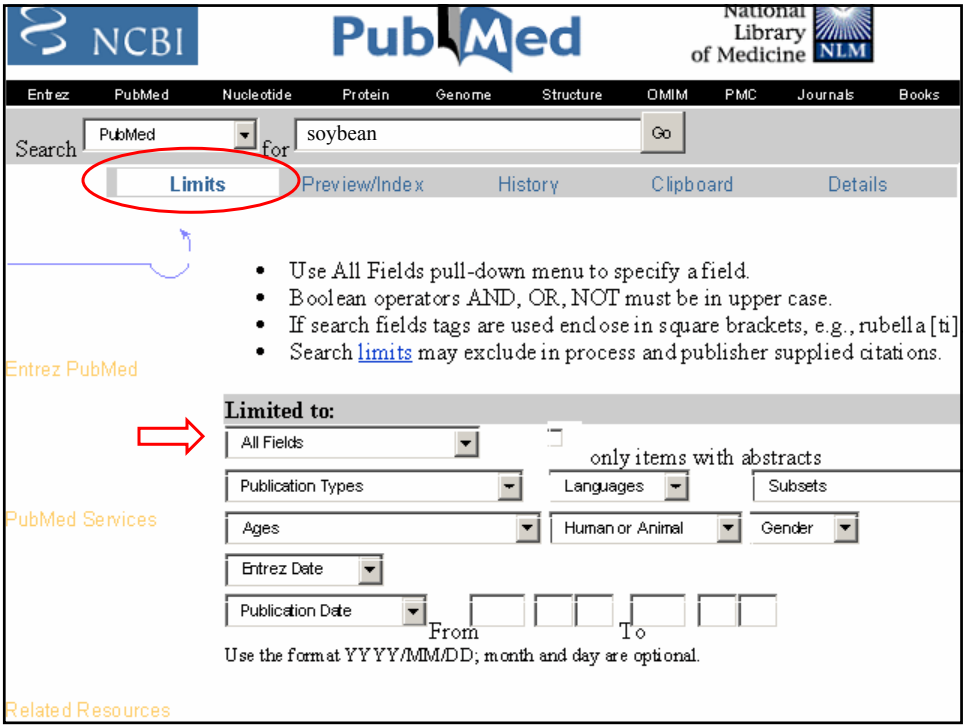
Entrez: Search and Retrieval System

http://www.ncbi.nlm.nih.gov/Entrez/

NCBI

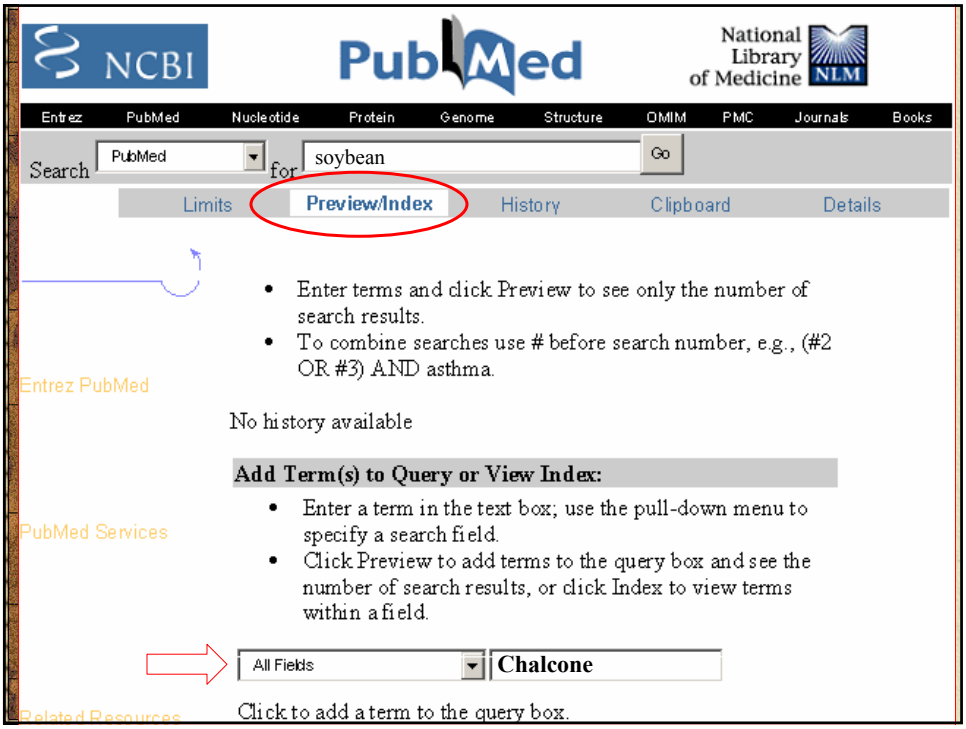

NCBI
PubMed
National Library of Medicine
NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Go
Limits Preview/Index History Clipboard Details
Entrez PubMed
Enter one or more search terms, or click Preview/Index for advanced searching.
Enter author names as smith jc. Initials are optional.
Enter journal titles in full or as MEDLINE abbreviations. Use the Journals Database to find journal titles.
PubMed Services
PubMed, a service of the National Library of Medicine, includes over 14 million citations for biomedical articles back to the 1950's. These citations are from MEDLINE and additional life science journals. PubMed includes links to many sites providing full text articles and

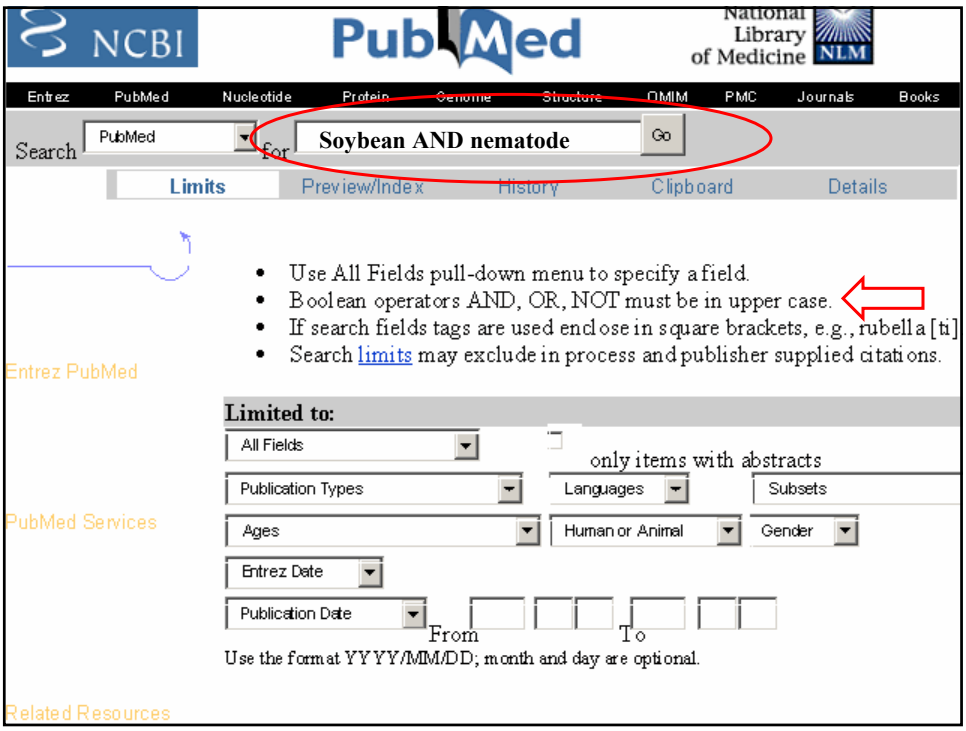

NCBI
PubMed
National Library of Medicine
NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for soybean Go
Limits Preview/Index History Clipboard Details
Entrez PubMed
Enter one or more search terms, or click Preview/Index for advanced searching.
Enter author names as smith jc. Initials are optional.
Enter journal titles in full or as MEDLINE abbreviations. Use the Journals Database to find journal titles.
PubMed Services
PubMed, a service of the National Library of Medicine, includes over 14 million citations for biomedical articles back to the 1950's. These citations are from MEDLINE and additional life science journals. PubMed includes links to many sites providing full text articles and



All Fields

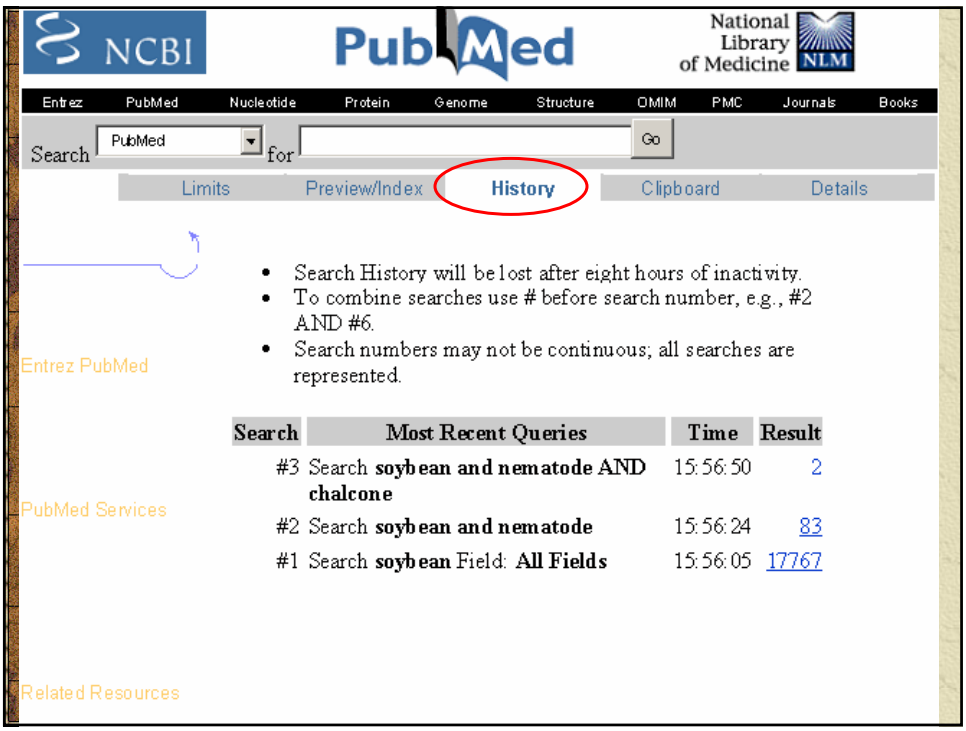
- Author
- Journal
- Issue


NCBI
PubMed
National Library of Medicine
NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for soybean Go
Limits Preview/Index History Clipboard Details
Entrez PubMed
Enter terms and click Preview to see only the number of search results.
To combine searches use # before search number, e.g., (#2 OR #3) AND asthma.
No history available
PubMed Services
Add Term(s) to Query or View Index:
Enter a term in the text box; use the pull-down menu to specify a search field.
Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.
All Fields
Chalcone
Related Resources
Click to add a term to the query box.


NCBI
PubMed
National Library of Medicine
NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Soybean AND nematode Go
Limits Preview/Index History Clipboard Details
Entrez PubMed
Use All Fields pull-down menu to specify a field.
Boolean operators AND, OR, NOT must be in upper case.
If search fields tags are used enclose in square brackets, e.g., rubella [ti]
Search limits may exclude in process and publisher supplied citations.
Limited to:
All Fields
only items with abstracts
Publication Types
Languages
Subsets
PubMed Services
Ages
Human or Animal
Gender
Entrez Date
Publication Date
From
To
Use the format YYYY/MM/DD; month and day are optional.
Related Resources

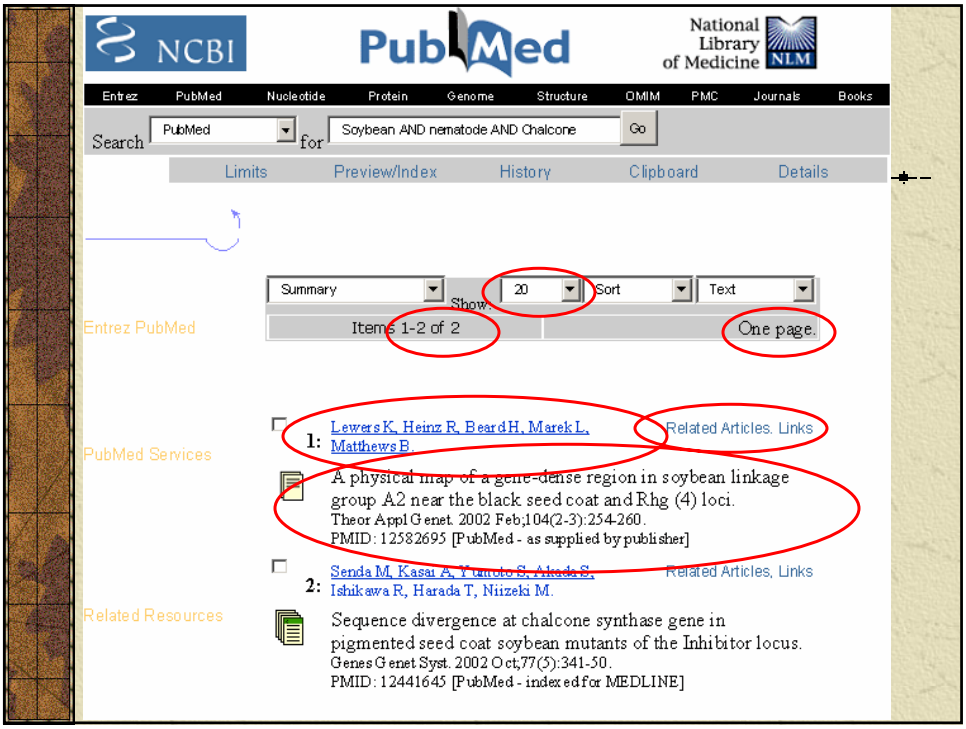
Keword searches

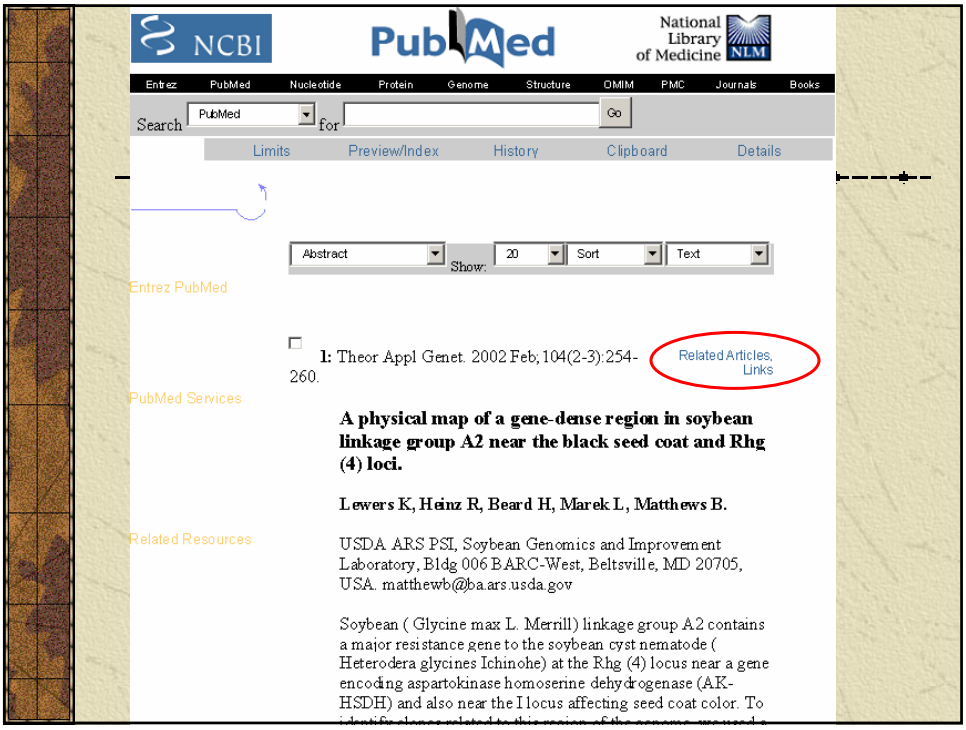
- Soybean
- Soybean nematode chalcone
- Soybean AND nematode NOT chalcone
  - Boolean operators: “AND” “OR” “NOT”
- “soybean cyst nematode”

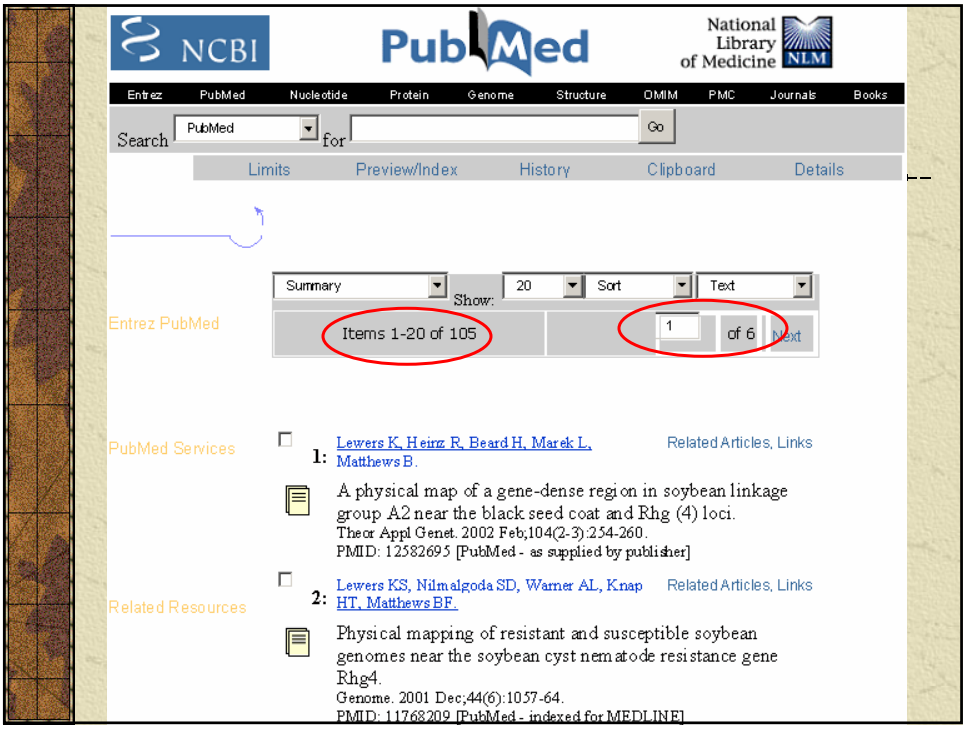



NCBI
PubMed
National Library of Medicine NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for chalcone soybean Go
Limits Preview/Index History Clipboard Details
Entrez PubMed
Summary Show: 20 Sort Text
Items 1-20 of 48
1 of 3 Next
PubMed Services
Related Resources
1: Lewers K, Heinz R, Beard H, Marek L, Matthews B. Related Articles, Links
A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg (4) loci.
Theor Appl Genet. 2002 Feb;104(2-3):254-260.
PMID: 12582695 [PubMed - as supplied by publisher]
2: Senda M, Jumonji A, Yumoto S, Ishikawa R, Harada T, Niizeki M, Akada S. Related Articles, Links
Analysis of the duplicated CHS1 gene related to the suppression of the seed coat pigmentation in yellow soybeans.
Theor Appl Genet. 2002 May;104(6-7):1086-1091. Epub 2002 Feb 20.
PMID: 12582616 [PubMed - as supplied by publisher]
3: Senda M, Kasai A, Yumoto S, Akada S, Ishikawa R, Harada T, Niizeki M. Related Articles, Links
Sequence divergence at chalcone synthase gene in pigmented seed coat soybean mutants of the Inhibitor locus.
Genes Genet Syst. 2002 Oct;77(5):341-50.
PMID: 12441645 [PubMed - indexed for MEDLINE]
4: Liu CJ, Blount JW, Steele CL, Dixon RA. Related Articles, Links
Bottlenecks for metabolic engineering of isoflavone glycoconjugates in Arabidopsis.
Proc Natl Acad Sci U S A. 2002 Oct 29;99(22):14578-83. Epub 2002 Oct 16.
PMID: 12384577 [PubMed - indexed for MEDLINE]


NCBI
PubMed
National Library of Medicine NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for chalcone soybean Go
Limits Preview/Index History Clipboard Details
Entrez PubMed
Summary Show: 20 Sort Text
Items 1-20 of 48
1 of 3 Next
PubMed Services
Related Resources
1: Lewers K, Heinz R, Beard H, Marek L, Matthews B. Related Articles, Links
A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg (4) loci.
Theor Appl Genet. 2002 Feb;104(2-3):254-260.
PMID: 12582695 [PubMed - as supplied by publisher]
2: Senda M, Jumonji A, Yumoto S, Ishikawa R, Harada T, Niizeki M, Akada S. Related Articles, Links
Analysis of the duplicated CHS1 gene related to the suppression of the seed coat pigmentation in yellow soybeans.
Theor Appl Genet. 2002 May;104(6-7):1086-1091. Epub 2002 Feb 20.
PMID: 12582616 [PubMed - as supplied by publisher]
3: Senda M, Kasai A, Yumoto S, Akada S, Ishikawa R, Harada T, Niizeki M. Related Articles, Links
Sequence divergence at chalcone synthase gene in pigmented seed coat soybean mutants of the Inhibitor locus.
Genes Genet Syst. 2002 Oct;77(5):341-50.
PMID: 12441645 [PubMed - indexed for MEDLINE]
4: Liu CJ, Blount JW, Steele CL, Dixon RA. Related Articles, Links
Bottlenecks for metabolic engineering of isoflavone glycoconjugates in Arabidopsis.
Proc Natl Acad Sci U S A. 2002 Oct 29;99(22):14578-83. Epub 2002 Oct 16.
PMID: 12384577 [PubMed - indexed for MEDLINE]


NCBI
PubMed
National Library of Medicine
NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Soybean AND nematode AND Chalcone Go
Limits Preview/Index History Clipboard Details
Summary Show: 20 Sort Text
Items 1-2 of 2
One page.
Entrez PubMed
PubMed Services
Related Resources
1: Lewers K, Heinz R, Beard H, Marek L, Matthews B.
Related Articles, Links
A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg (4) loci.
Theor Appl Genet. 2002 Feb;104(2-3):254-260.
PMID: 12582695 [PubMed - as supplied by publisher]
2: Senda M, Kasai A, Yamada S, Akada S, Ishikawa R, Harada T, Niizeki M.
Related Articles, Links
Sequence divergence at chalcone synthase gene in pigmented seed coat soybean mutants of the Inhibitor locus.
Genes Genet Syst. 2002 Oct;77(5):341-50.
PMID: 12441645 [PubMed - indexed for MEDLINE]


NCBI
PubMed
National Library of Medicine
NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Go
Limits Preview/Index History Clipboard Details
Abstract Show: 20 Sort Text
Entrez PubMed
PubMed Services
Related Resources
1: Theor Appl Genet. 2002 Feb;104(2-3):254-260.
Related Articles, Links
A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg (4) loci.
Lewers K, Heinz R, Beard H, Marek L, Matthews B.
USDA ARS PSI, Soybean Genomics and Improvement Laboratory, Bldg 006 BARC-West, Beltsville, MD 20705, USA. matthewb@ba.ars.usda.gov
Soybean (Glycine max L. Merrill) linkage group A2 contains a major resistance gene to the soybean cyst nematode (Heterodera glycines Ichinohe) at the Rhg (4) locus near a gene encoding aspartokinase homoserine dehydrogenase (AK-HSDH) and also near the I locus affecting seed coat color. To

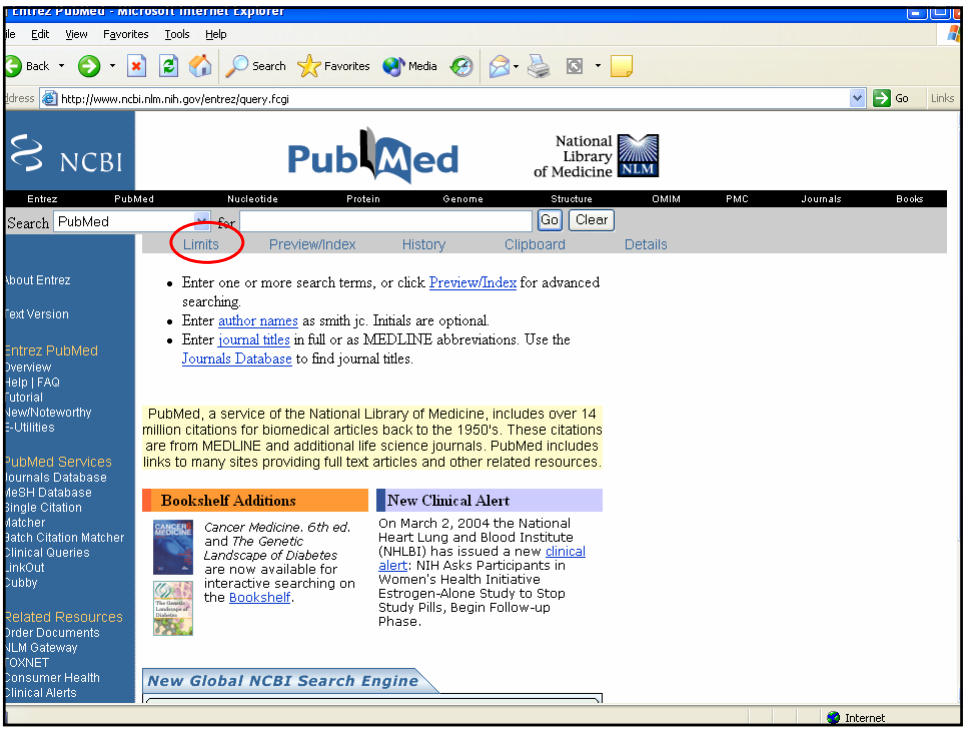


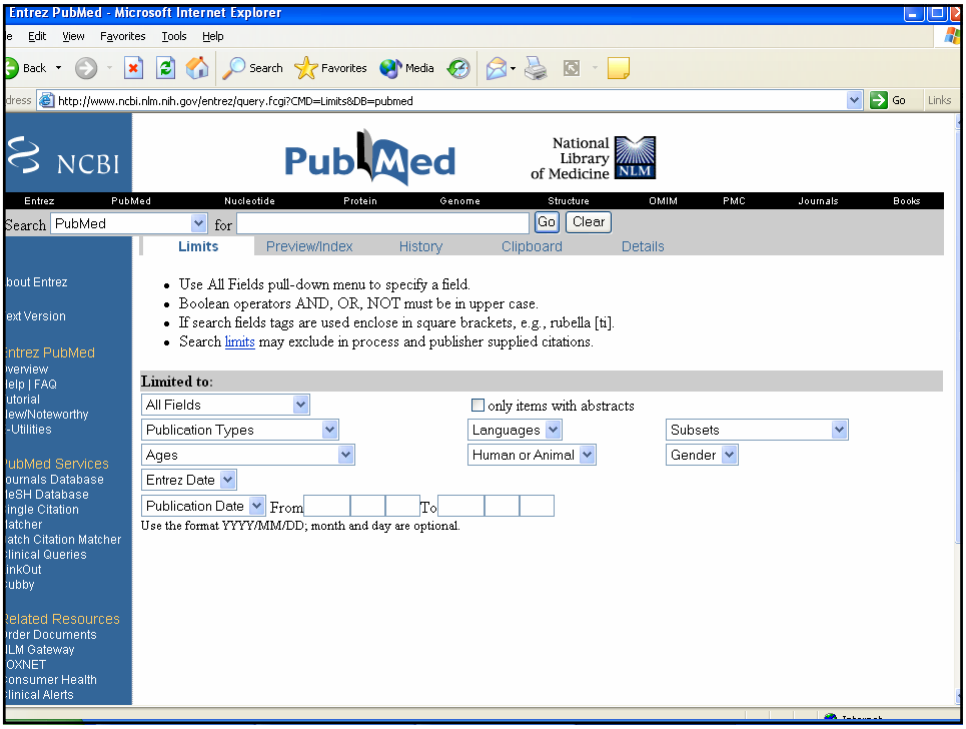
Other searches

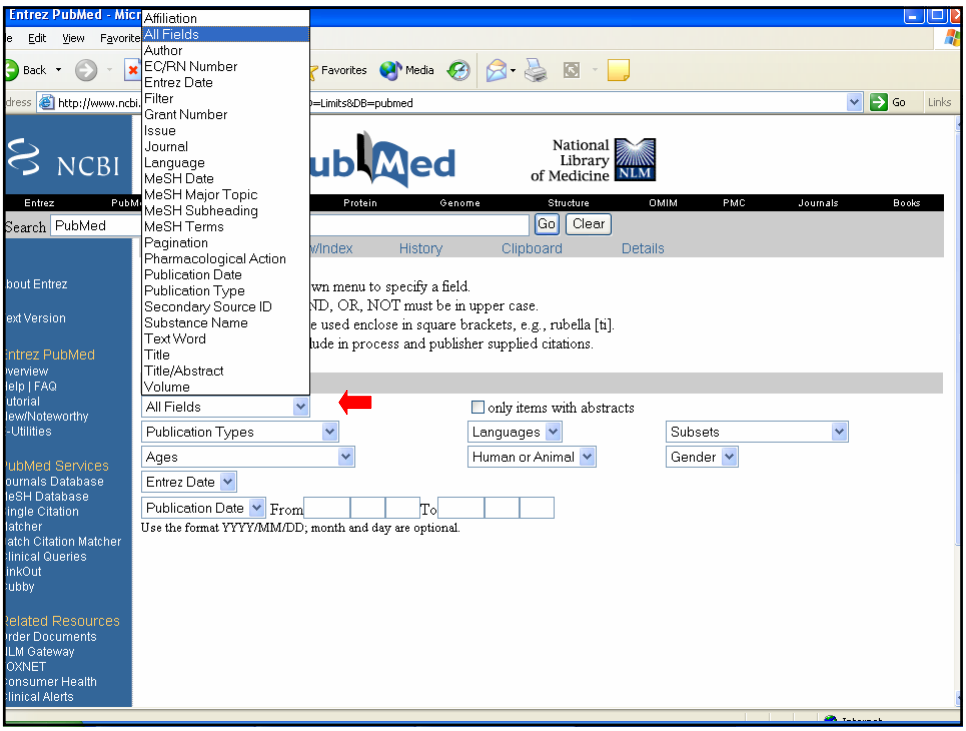
- Author's name-query
  - Find other papers by the author that may be related to the topic
  - Find paper you remember author but not source
- Field-restricted query
  - Narrows search
  - Dates, address, journal

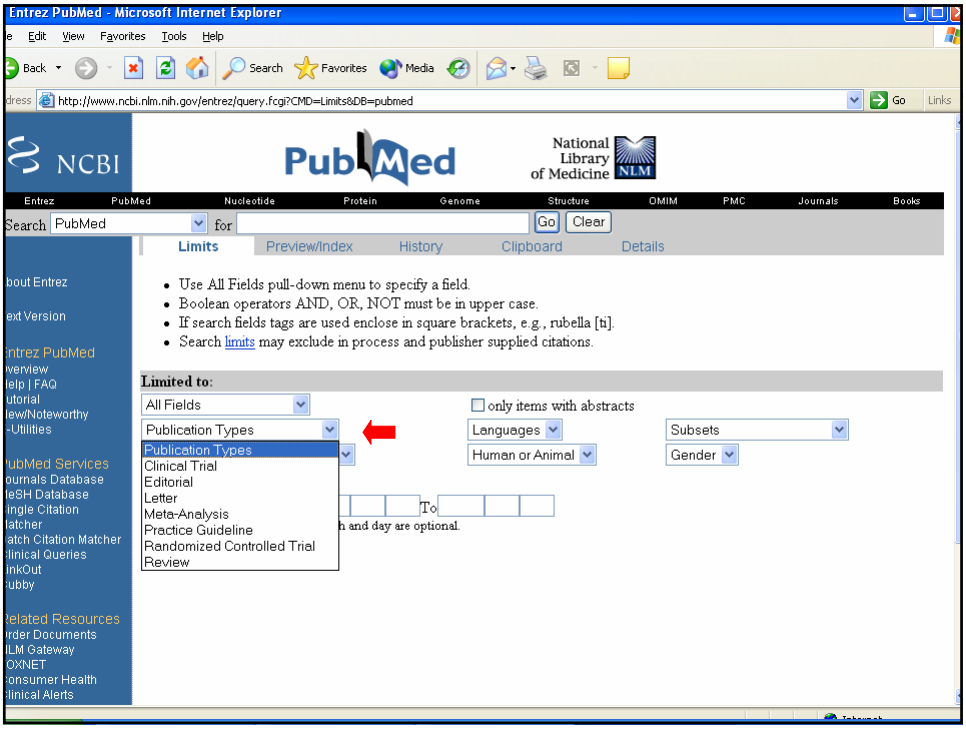
Limits

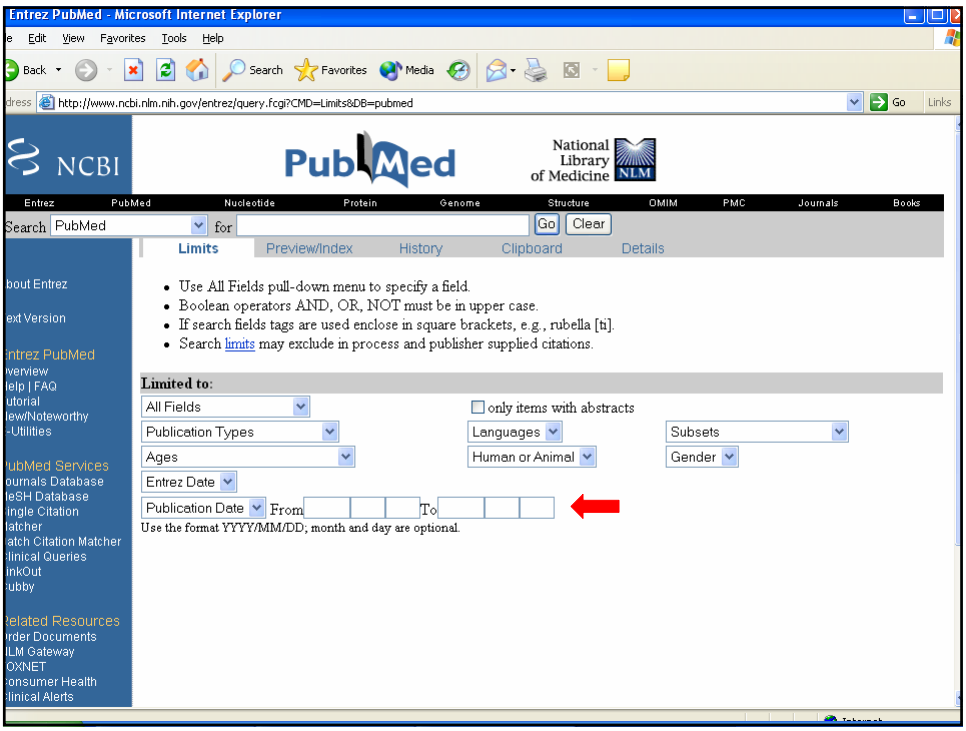
- Narrow the focus of a search
- Publication types

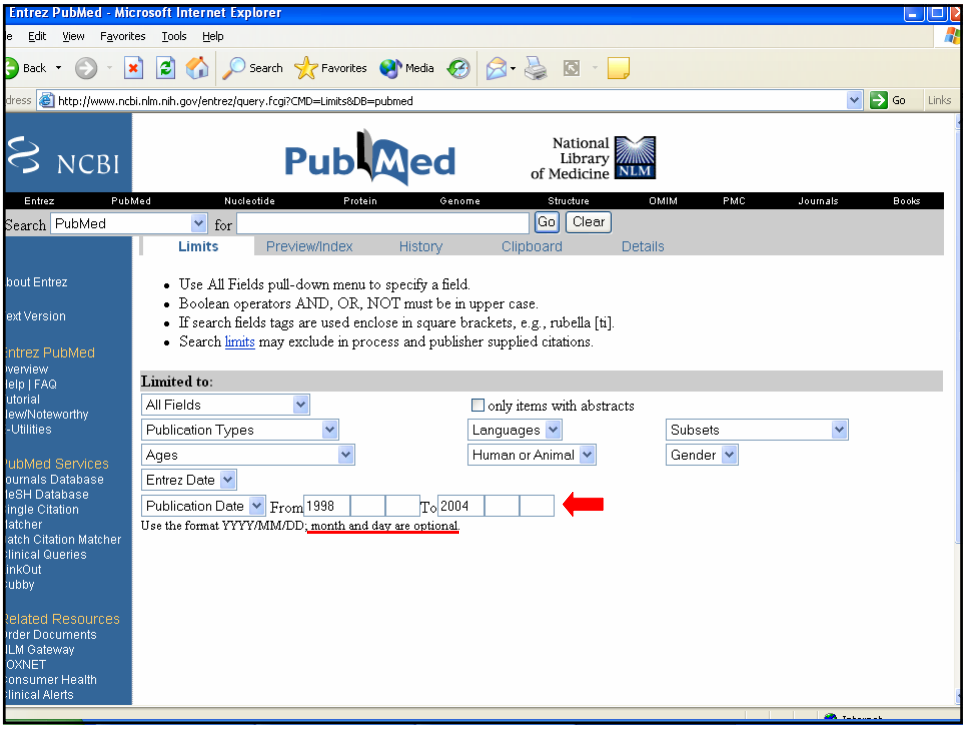


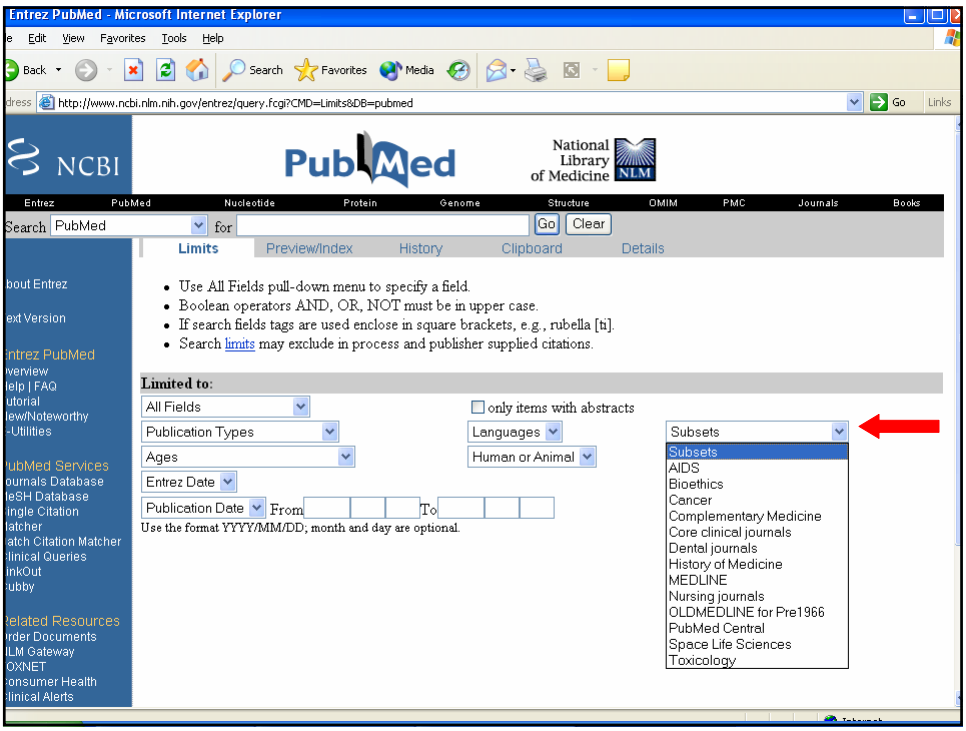

Entrez PubMed - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=pubmed
Go Links
NCBI
PubMed
National Library of Medicine NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Go Clear
Limits Preview/Index History Clipboard Details
About Entrez
Text Version
Entrez PubMed
Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities
PubMed Services
Journals Database
MeSH Database
Single Citation
Matcher
Batch Citation Matcher
Clinical Queries
LinkOut
Cubby
Related Resources
Order Documents
NLM Gateway
OXNET
Consumer Health
Clinical Alerts
Use All Fields pull-down menu to specify a field.
Boolean operators AND, OR, NOT must be in upper case.
If search fields tags are used enclose in square brackets, e.g., rubella [ti].
Search limits may exclude in process and publisher supplied citations.
Limited to:
All Fields
only items with abstracts
Publication Types
Languages
Subsets
Ages
Human or Animal
Gender
Entrez Date
Publication Date From To
Use the format YYYY/MM/DD; month and day are optional.


Affiliation
All Fields
Author
EC/RN Number
Entrez Date
Filter
Grant Number
Issue
Journal
Language
MeSH Date
MeSH Major Topic
MeSH Subheading
MeSH Terms
Pagination
Pharmacological Action
Publication Date
Publication Type
Secondary Source ID
Substance Name
Text Word
Title
Title/Abstract
Volume
All Fields
only items with abstracts
Publication Types
Languages
Subsets
Ages
Human or Animal
Gender
Entrez Date
Publication Date From To
Use the format YYYY/MM/DD; month and day are optional.


Entrez PubMed - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=pubmed
Go Links
NCBI
PubMed
National Library of Medicine NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Go Clear
Limits Preview/Index History Clipboard Details
bout Entrez
ext Version
ntrez PubMed
verview
elp | FAQ
utorial
ew/Noteworthy
-Utilities
ubMed Services
ournals Database
eSH Database
ingle Citation
latcher
atch Citation Matcher
linical Queries
inkOut
ubby
elated Resources
rder Documents
LM Gateway
OXNET
onsumer Health
linical Alerts
Use All Fields pull-down menu to specify a field.
Boolean operators AND, OR, NOT must be in upper case.
If search fields tags are used enclose in square brackets, e.g., rubella [ti].
Search limits may exclude in process and publisher supplied citations.
Limited to:
All Fields
only items with abstracts
Publication Types
Languages
Subsets
Publication Types
Clinical Trial
Editorial
Letter
Meta-Analysis
Practice Guideline
Randomized Controlled Trial
Review
Human or Animal
Gender
To
and day are optional.


Entrez PubMed - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=pubmed
Go Links
NCBI
PubMed
National Library of Medicine NLM
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Go Clear
Limits Preview/Index History Clipboard Details
bout Entrez
ext Version
ntrez PubMed
verview
elp | FAQ
utorial
ew/Noteworthy
-Utilities
ubMed Services
ournals Database
eSH Database
ingle Citation
latcher
atch Citation Matcher
linical Queries
inkOut
ubby
elated Resources
rder Documents
LM Gateway
OXNET
onsumer Health
linical Alerts
Use All Fields pull-down menu to specify a field.
Boolean operators AND, OR, NOT must be in upper case.
If search fields tags are used enclose in square brackets, e.g., rubella [ti].
Search limits may exclude in process and publisher supplied citations.
Limited to:
All Fields
only items with abstracts
Publication Types
Languages
Subsets
Ages
Human or Animal
Gender
Entrez Date
Publication Date From To
Use the format YYYY/MM/DD; month and day are optional.


Entrez PubMed - Microsoft Internet Explorer
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=pubmed
NCBI
PubMed
National Library of Medicine
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Go Clear
Limits Preview/Index History Clipboard Details
Use All Fields pull-down menu to specify a field.
Boolean operators AND, OR, NOT must be in upper case.
If search fields tags are used enclose in square brackets, e.g., rubella [ti].
Search limits may exclude in process and publisher supplied citations.
Limited to:
All Fields
only items with abstracts
Publication Types
Languages
Subsets
Ages
Human or Animal
Gender
Entrez Date
Publication Date From 1998 To 2004
Use the format YYYY/MM/DD; month and day are optional.


Entrez PubMed - Microsoft Internet Explorer
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=pubmed
NCBI
PubMed
National Library of Medicine
Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Go Clear
Limits Preview/Index History Clipboard Details
Use All Fields pull-down menu to specify a field.
Boolean operators AND, OR, NOT must be in upper case.
If search fields tags are used enclose in square brackets, e.g., rubella [ti].
Search limits may exclude in process and publisher supplied citations.
Limited to:
All Fields
only items with abstracts
Publication Types
Languages
Subsets
Ages
Human or Animal
Entrez Date
Publication Date From To
Use the format YYYY/MM/DD; month and day are optional.
Subsets
AIDS
Bioethics
Cancer
Complementary Medicine
Core clinical journals
Dental journals
History of Medicine
MEDLINE
Nursing journals
OLDMEDLINE for Pre1966
PubMed Central
Space Life Sciences
Toxicology

Tips

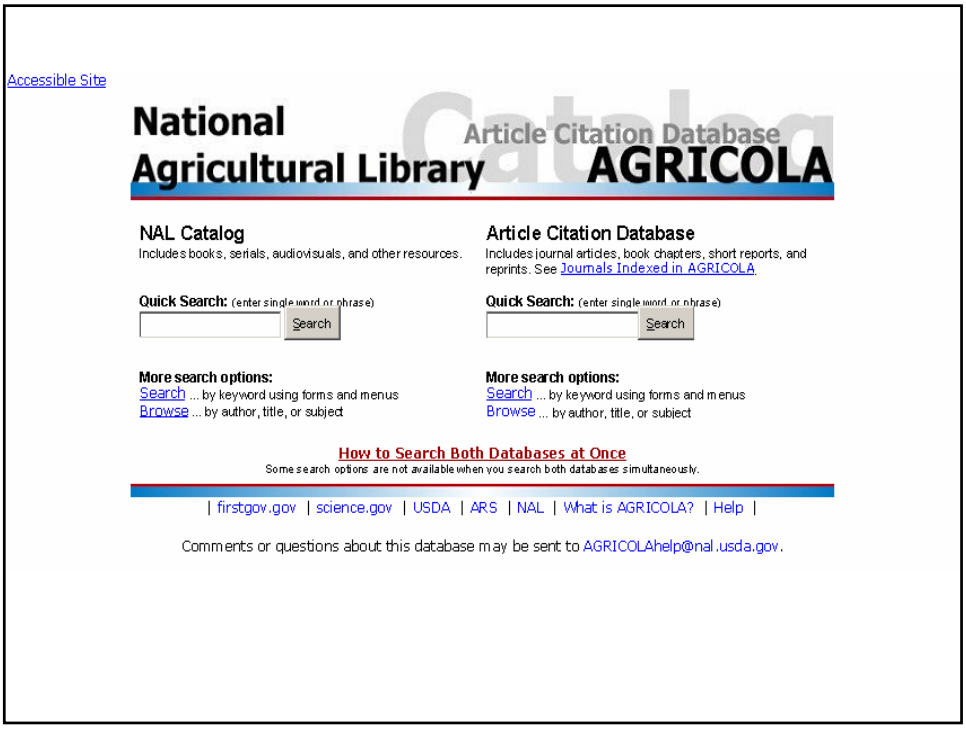
- Quoted queries behave as a single word
  - Example …"Soybean cyst nematode"
- Add initials to author queries
  - Example…Matthews B
- If you don't find what you want, don't give up
  - Try a different key word
  - Indexing terminology is important
- Try "Related papers" button
- No papers before 1965 are here
- No abstracts for papers before 1976 are here

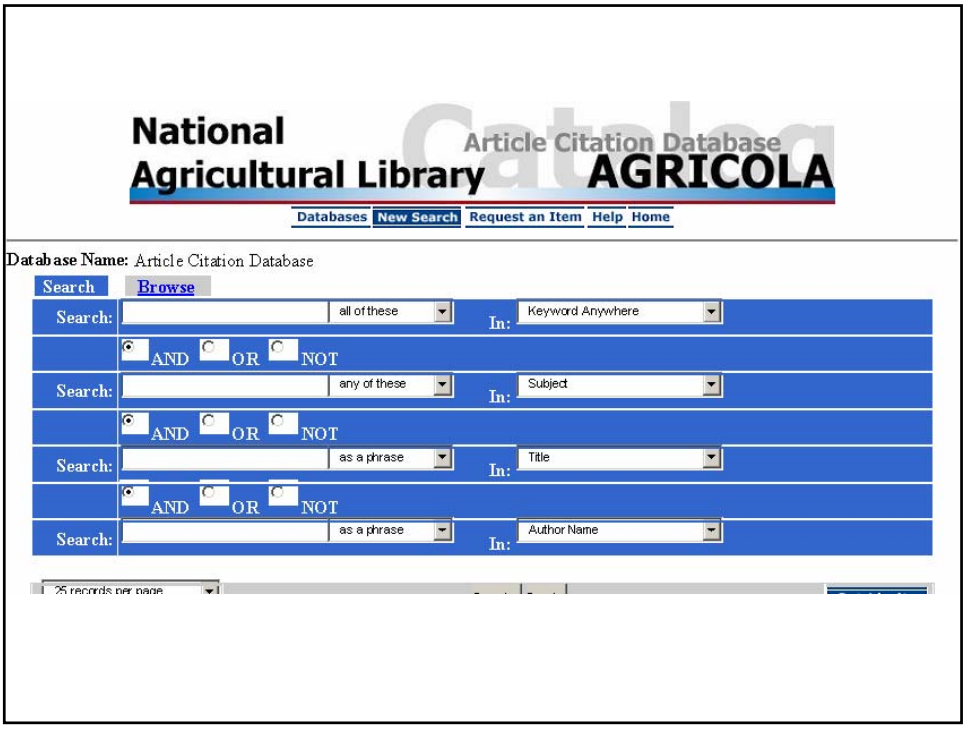
Agricola

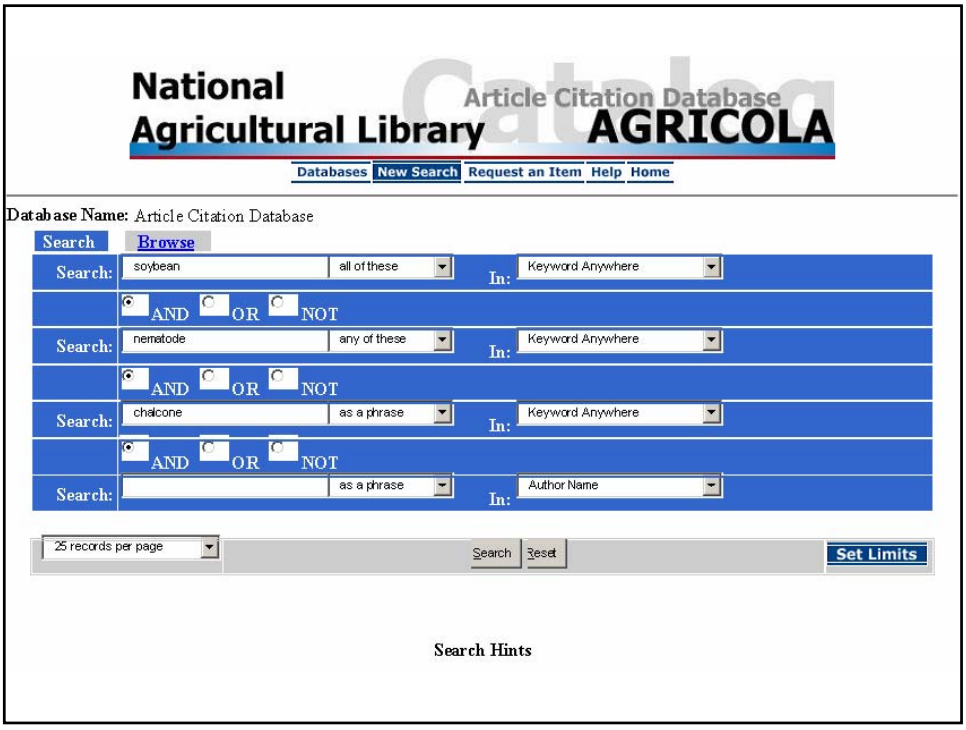
- **AGRI**Cultural **O**n **L**ine **A**ccess
- Bibliographic database of citations to agricultural literature
- Two databases
  - NAL catalog
    - Books, serials, audio-visuals
  - Article Citation Index
    - Journal articles
- U.S. Department of Agriculture
- National Agricultural Library
- http://agricola.nal.usda.gov/

PubMed vs Agricola

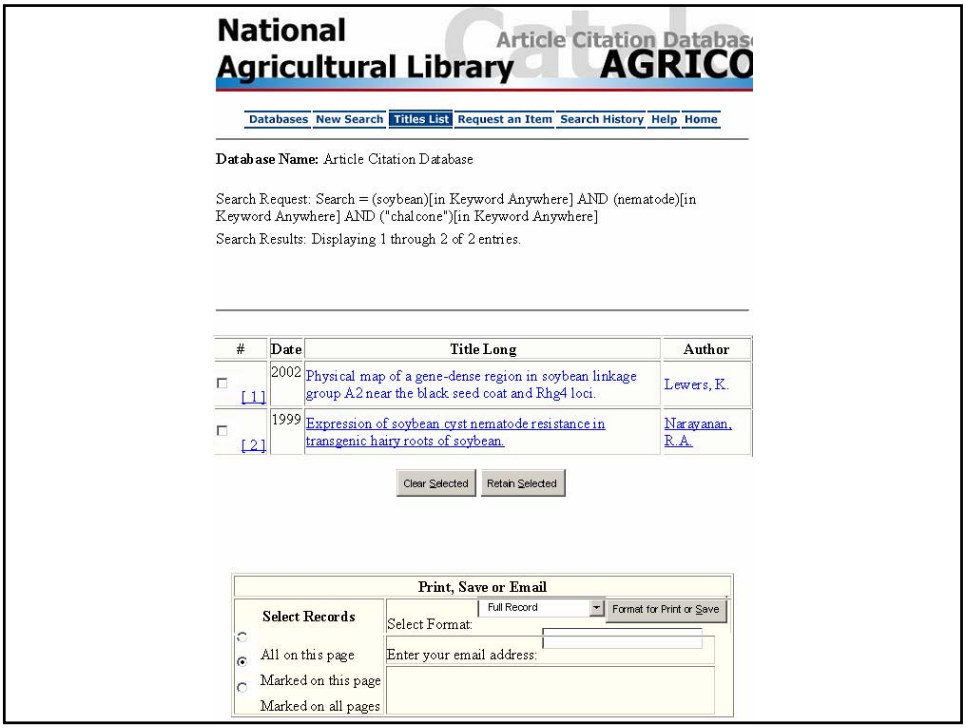
- Substantial overlap of journals indexed
- PubMed
  - Broader molecular, medicine, biology
  - Extensive
  - No basic agronomy
- Agricola
  - Agriculture-related
  - Narrower focus
  - Basic agronomy found here







Search Hints



What we covered today

- What is bioinformatics
- Databases
- Literature searches